DATASET BRIEF

The proteome profile of embryogenic cell suspensions of *Coffea arabica* L.

Nádia A. Campos[1], Luciano V. Paiva[2], Bart Panis[3] and Sebastien C. Carpentier[1,4]

[1] Biosystems Department, KULeuven, Leuven, Belgium
[2] Chemical Department, Federal University of Lavras, Minas Gerais, Brazil
[3] Bioversity International, Leuven, Belgium
[4] SYBIOMA: Facility for SYstems BIOlogy based MAss spectrometry, Leuven, Belgium

Somatic embryogenesis, is a process by which new viable embryos are produced from somatic tissues. Somatic embryogenesis is not only a useful biotechnological tool for the massive clonal propagation and genetic engineering but it also allows to obtain fundamental knowledge about the molecular changes that take place during embryogenesis. We present the proteome profile of two embryogenic cell suspensions. We identified 1052 non-redundant proteins. We present their known GO annotations and show two protein networks sharing the GO annotations related to stress and embryogenic capacity via the free program Cytoscape. To our knowledge these results give the first high-throughput proteome description of embryogenic cell suspensions and provide new information about somatic embryos for the whole plant community. The published proteome is a first step toward understanding somatic embryogenesis in coffee and toward a better annotation of proteins in an important non-model crop. All data are available via ProteomeXchange with identifier PXD002963.

Received: October 14, 2015
Revised: December 3, 2015
Accepted: January 11, 2016

Keywords:
Annotation / Coffee / Orphan species / Plant proteomics / Somatic embryogenesis

Additional supporting information may be found in the online version of this article at the publisher's web-site

Coffee is one of the most important commodities worldwide and has a great economic impact in producing countries. Although 130 different species belonging to the *Coffea* gender have been described, only two of them are commercially exploited: *Coffea arabica* and *Coffea canephora*. *Coffea arabica* is responsible for 61% of the world production [1]. However, due to the narrow genetic back ground and time to complete the life cycle, classical genetic breeding is time consuming [2, 3]. Several biotechnological tools are already applied in coffee breeding. Somatic embryogenesis (SE) is a process in which new viable embryos are produced from somatic tissues. It is one of the most promising mass propagation processes [2, 4].

Correspondence: Dr. Sebastien Carpentier, Department of Biosystems, KULeuven, Willem Decroylaan 42 bus 2455-3001 Leuven, Belgium Leuven, 3000, Belgium
E-mail: Sebastien.Carpentier@kuleuven.be

Abbreviations: BP, biological process; **CC**, cellular component; **GO**, gene ontology; **MF**, molecular function; **PGDH**, phosphoglycerate dehydrogenase; **SE**, somatic embryogenesis

A better understanding of the molecular basis related to SE will give insight into the process of embryo formation and totipotency and will allow the development of new in vitro culture strategies for the propagation and genetic improvement of elite cultivars [4, 5]. High-throughput proteomics in crops and coffee in particular is mostly limited to 2D gel-based techniques. Although really useful and still the most common technique for plants, 2DE is limited in its throughput and gel-free techniques allow to go a step further [6–8]. To improve the knowledge about coffee SE, we present the first high-throughput proteome profile of embryogenic cell suspensions (1052 protein identifications, FDR 0.3%).

The leaves of *Coffea arabica* cultivar Catuaí were collected from plants kept in a greenhouse at the Federal University of Lavras and cultivated in vitro following the protocol described in Silva et al. [10] to induce embryogenesis. Six-month-old calli were used to initiate cell suspensions that were cultivated according to Silva et al. [10]. Three months later, two different

Colour Online: See the article online to view Figs. 1–3 in colour.

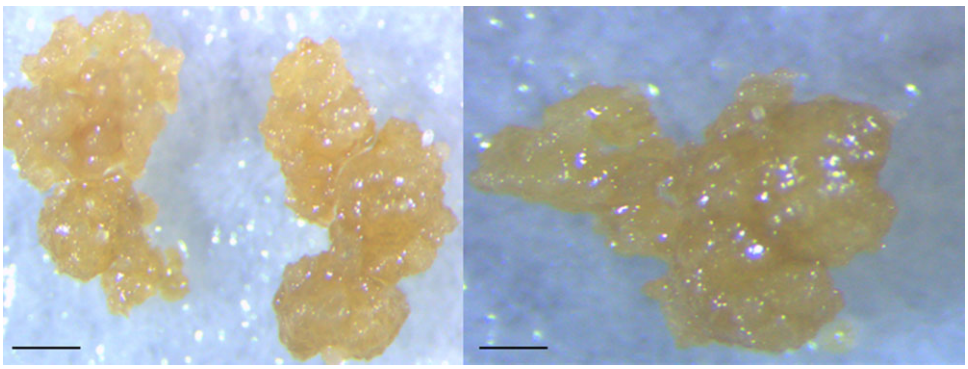

Figure 1. Cell suspensions. Left cell line 1, right cell line 2. Bars = 0.5 cm.

cell suspensions (Fig. 1) considered as embryogenic based in morphological observations [9] were used to extract the proteins. The embryo regeneration capacity was estimated based on two samples with the same morphological characteristics (Fig. 1). The embryo regeneration capacity had an average of 26 embryos/g of cell ($n = 6$). Cells were divided in 21 (12 from line 1 and nine from line 2) aliquots of 30 mg FW. Protein extraction took place by the phenol extraction protocol reported by Carpentier et al. [11] and adapted by Buts et al. [12].

The digested samples (1 μg /5 μL) were injected and separated on an Ultimate 3000 UPLC system (Dionex, ThermoScientific) equipped with a Acclaim PepMap100 pre column (C18 3 μm–100 Å, Thermo Scientific) and an C18 PepMap RSLC (2 μm, 50 μm_15cm,ThermoScientific) using a linear gradient (300 μL/min) of 0–4% buffer B (80% ACN, 0.08% FA) in 3 min, 4–10% B in 12 min, 10–35% in 20 min, 35–65% in 5 min, 65–95% in 1 min 95% for 10 min, 95–5% in 1 min, 5% 10 min. The Q Exactive Orbitrap mass spectrometer (Thermo Scientific, USA) was operated in positive ion mode with a nanospray voltage of 1.5 kV and a source temperature of 250°C. Proteo Mass LTQ/FT-Hybrid ESI Pos. Mode Cal Mix (MS CAL5-1EASUPELCO, Sigma-Aldrich) was used as an external calibrant and the lock mass 445.12003 as an internal calibrant. The instrument was operated in data-dependent acquisition (DDA) mode with a survey MS scan at a resolution of 70 000 (fw hm at m/z 200) for the mass range of m/z 400–1600 for precursor ions, followed by MS/MS scans of the top ten most intense peaks with +2, +3, +4, and +5 charged ions above a threshold ion count of 16 000 at 17 500 resolution using normalized collision energy (NCE) of 25 eV with an isolation window of 3.0 m/z and dynamic exclusion of 10 s. All data were acquired with Xcalibur 3.0.63 software (ThermoScientific). For identification, all raw data were converted into mgf.files by Proteome Discover version 1.4 (Thermo Scientific) and processed using MASCOT version 2.2.06 (Matrix Science) against the Uniprot coffea database (25 602 proteins). The parameters used to search at MASCOT were: parent tolerance of 10 PPM, fragment tolerance of 0.02 Da, variable modification oxidation of M, fixed modification with carbamidomethyl C and up to one missed cleavage for trypsin. Results from MASCOT were imported to Scaffold version 3.6.3. The parameters used in Scaffold for protein identification was to retain proteins containing at least one identified peptide with confidence level 95%. The resulting false discovery rate on protein level was 0.3%. In total we identified 1052 proteins (Supporting Information Table 1). The raw data have been deposited to the ProteomeXchange consortium [13] via PRoteomics IDEntifications (PRIDE) [14] partner repository with identifier PXD002963 and 10.6019/PXD002963. Annotation and cross species annotation is a big challenge for coffee and other non-model crops. Seven hundred forty-six identified proteins were linked by Uniprot to at least one GO annotation, resulting in 2032 GO annotations: 1109 belonging to molecular function term (MF), 577 to biological process (BP), and 346 cellular component (CC) (Supporting Information Table 1). Three hundred and six genes/proteins are without GO annotation and 279 are still without any GO annotation or protein name, illustrating the potential of this dataset to annotate genomes. For all identified proteins, the GOs were retrieved from Uniprot and the protein/GO interactions were loaded into Cytoscape to analyze the similarities between the identified proteins. Such a network enables a better annotation of certain genes/paralogs and to gain better insights into the important biological processes and molecular functions that are associated to somatic embryos.

It is generally accepted that the process of somatic embryogenesis (SE) is correlated to biochemical changes leading to reprograming of gene expression, resulting in cell division [15]. Many of these genes are related to stress [16]. Our results show 24 proteins linked to three different gene ontology terms related to stress (GO 0006950, GO 0006979, and GO 0034599) (Fig. 2). One of these proteins (A0A068UMU7) originally without protein name is correlated to two more GOs (GO 0020037 – heme binding and GO 0004601 – peroxidase activity). A blast shows a high similarity (>90%) with ascorbate peroxidase in many other organisms (e.g., *Nicotiana tabacum, Solanum lycopersicum, Capsicum annuum*). Ascorbate peroxidase plays an important role in heat stress response, oxidative stress, and stress tolerance and is known to be part of the reactions for scavenging reactive oxygen species [17, 18].

As many of the proteins identified are not annotated for coffee, we used Blast2Go to check the potential embryogenic proteins via cross-species annotation and added the new annotations to our dataset. Next to A0A068UMU7, Blast2Go

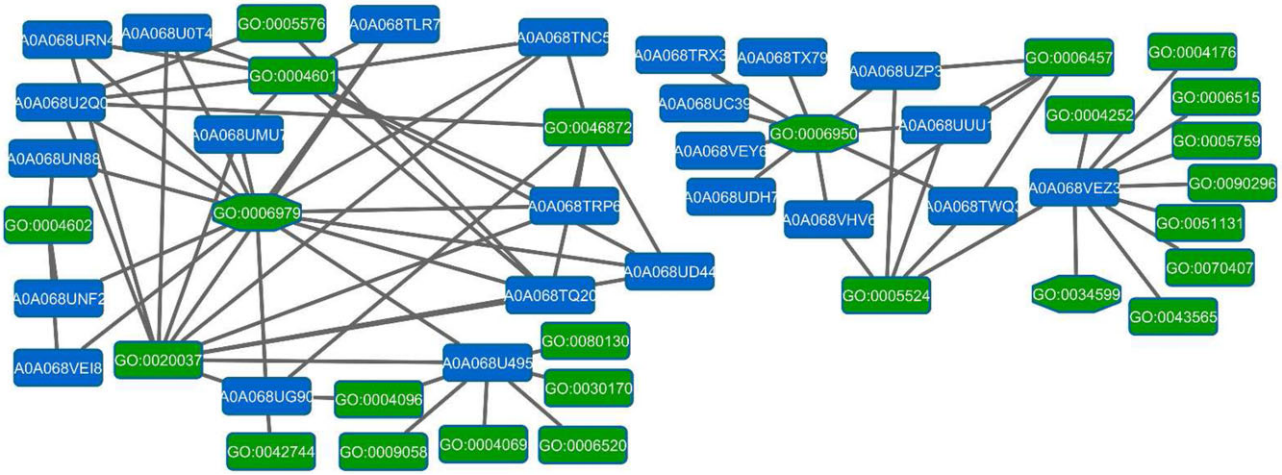

Figure 2. Cytoscape GO-Protein Networks related to stress. Blue: identified proteins. Green: GO annotated via Uniprot. Octagon: initiator of the networks, GO directly linked to stress.

returned four additional proteins (A0A068TM02, A0A68URE9, A0A068VGZ4, and A0A068TL38) that are linked to GO numbers related to embryogenic processes. These proteins are described below and illustrated on Fig. 3. Protein A0A068TM02 was already associated to four GOs (GO 0004617, GO 0006564, GO 0016597, and GO 0051287). Two of them were described as phosphoglycerate dehydrogenase activity and L-serine biosynthetic process enzyme and linked via Blast2GO to embryogenesis. According to Toujani et al. [19], phosphoglycerate dehydrogenase (PGDH) is essential for embryo and pollen development in Arabidopsis, besides the importance in the pathway of serine metabolism. Protein A0A068URE9 has more than 60% similarity with a chitinase identified in rice and maize and has activity in carbohydrate metabolism [20]. A protein with identity A0A068VGZ4 has more than 80% similarity with a protein identified in Arabidopsis as an acetyl-CoA acetyltransferase. A study of Arabidopsis mutants showed that this protein is strongly required for embryogenesis and normal male gamete transmission [21]. The protein A0A068TL38, has 60% of similarity with a cold shock protein and can be an important key to explain somatic embryogenesis. Among

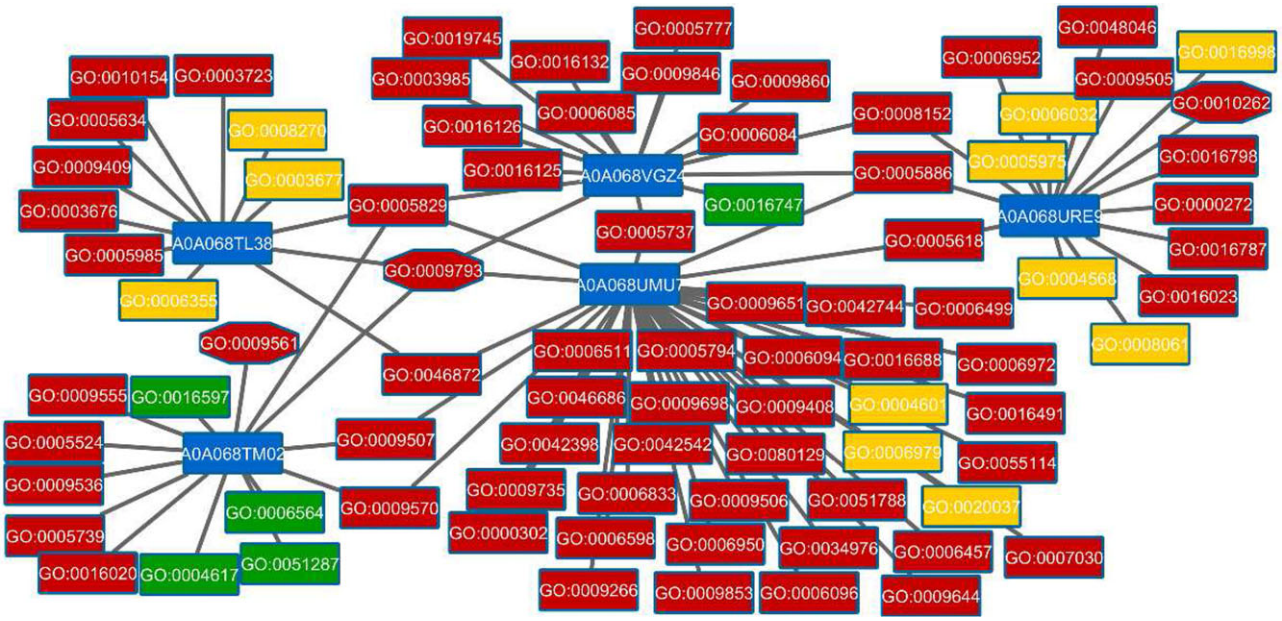

Figure 3. Cytoscape GO-Protein Network related to embryogenic capacity. Blue: identified proteins. Green: GO uniquely annotated via Uniprot. Red: GO uniquely retrieved via cross species Blast2GO. Yellow: annotated via Uniprot and Blast2GO. Octagon: initiator of the network, GO directly linked to embryogenesis.

the four cold shock proteins known in Arabidopsis, one is characterized for affecting flowering time and seed development [22]. In the same context, Yang et al. [23] showed the importance of cold shock domain protein in late stages of embryo development.

Studies in different stages of somatic embryogenesis in coffee [24], embryogenic callus in sugarcane [25] and maize [26] proved that many proteins linked to the energy production are also important for embryogenesis. Those studies correlate enolase, 2-3-biphosphoglycerate-independent phosphoglycerate mutase, glyceraldehyde-3-phosphate dehydrogenase, and aldolases (cytoplasmic aldolase, fructose bisphosphate aldolase) to somatic embryogenesis. This is understandable since more energy is needed to induce intense cell division leading to new embryos [27]. We were able to identify different aldolases such as fructose bisphosphate aldolase (GO 0004332/ GO 0006096) and 4-hydroxy-4-methyl-2-oxoglutarate aldolase that represent four different GOs (GO 0046872, GO0008428, GO0047443, and GO 0051252) (Supporting Information Table 1).

The proteins described above can be good candidates to be markers for embryogenic cells and can play a role in guiding the improvement of somatic embryo production. Our results present the first report of proteins extracted from embryogenic calli of *Coffea arabica* using a shotgun approach and will facilitate the characterization of somatic embryogenesis in this important crop. A next step is to analyze comparatively embryogenic and non-embryogenic suspensions.

The authors are grateful to CNPq (National Counsel of Technological and Scientific Development/Brazil) for the scholarship of Dr. Nadia Campos, to Msc. Natalia Chagas for collecting the leaves in Lavras and to Kusay Arat for technical support at Sybioma.

The authors have declared no conflict of interest.

References

[1] Van der Vossen, H., Bertrand, B., Charrier, A., Next generation variety development for sustainable production of arabica coffee (Coffea arabica L.): a review. *Euphytica* 2015, *204*, 243–256.

[2] Santana-Buzzy, N., Rojas-Herrera, R., Galaz-Ávalos, R. M., Ku-Cauich, J. R. et al., Advances in coffee tissue culture and its practical applications. *Vitr. Cell. Dev. Biol. - Plant* 2007, *43*, 507–520.

[3] Hendre, P. S., Aggarwal, R. K., Development of genic and genomic SSR markers of Robusta coffee (Coffea canephora Pierre Ex A. Froehner). *PLoS One* 2014, *9*, e113661.

[4] Marsoni, M., Bracale, M., Espen, L., Prinsi, B. et al., Proteomic analysis of somatic embryogenesis in Vitis vinifera. *Plant Cell Rep.* 2008, *27*, 347–356.

[5] Guzmán-García, E., Sánchez-Romero, C., Panis, B., Carpentier, S. C., The use of 2D-DIGE to understand the regeneration of somatic embryos in avocado. *Proteomics* 2013, *13*, 3498–3507.

[6] Carpentier, S. C., America, T., *Plant Proteomics* 2014, *1072*, 333–346.

[7] Vanhove, A.-C., Vermaelen, W., Swennen, R., Carpentier, S. C., A look behind the screens: characterization of the HSP70 family during osmotic stress in a non-model crop. *J. Proteomics* 2015, *119*, 10–20.

[8] Vanhove, A.-C., Screening the banana biodiversity for drought tolerance: can an in vitro growth model and proteomics be used as a tool to discover tolerant varieties and understand homeostasis. *Front. Plant Sci.* 2012, *3*, 1–10.

[9] Padua, M. S., Paiva, L. V., Silva, L. C., Livramento, K. G. Do et al., Morphological characteristics and cell viability of coffee plants calli. *Ciência Rural* 2014, *44*, 660–665.

[10] Silva, A. T., Barduche, D., do Livramento, K. G., Ligterink, W., Paiva, L. V., Characterization of a putative Serk-like ortholog in embryogenic cell suspension cultures of Coffea arabica L. *Plant Mol. Biol. Report.* 2014, *32*, 176–184.

[11] Carpentier, S. C., Witters, E., Laukens, K., Deckers, P. et al., Preparation of protein extracts from recalcitrant plant tissues: an evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics* 2005, *5*, 2497–2507.

[12] Buts, K., Michielssens, S., Hertog, M.L.A.T.M., Hayakawa, E. et al., Improving the identification rate of data independent label-free quantitative proteomics experiments on non-model crops: a case study on apple fruit. *J. Proteomics* 2014, *105*, 31–45.

[13] Vizcaíno, J. A, Deutsch, E. W., Wang, R., Csordas, A. et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* 2014, *32*, 223–226.

[14] Vizcaino, J. A., Cote, R. G., Csordas, A., Dianes, J. A. et al., The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 2013, *41*, D1063–D1069.

[15] Fehér, A., Pasternak, T. P., Dudits, D., Transition of somatic plant cells to an embryogenic state. *Plant Cell. Tissue Organ Cult.* 2003, *74*, 201–228.

[16] Fehér, A., Somatic embryogenesis — Stress-induced remodeling of plant cell fate. *Biochim. Biophys. Acta - Gene Regul. Mech.* 2015, *1849*, 385–402.

[17] Panchuk, I., Volkov, R., Schöffl, F., Heat stress-and heat shock transcription factor-dependent expression and activity of ascorbate peroxidase in Arabidopsis. *Plant Physiol.* 2002, *129*, 838–853.

[18] Maruta, T., Tanouchi, A., Tamoi, M., Yabuta, Y. et al., Arabidopsis chloroplastic ascorbate peroxidase isoenzymes play a dual role in photoprotection and gene regulation under photooxidative stress. *Plant Cell Physiol.* 2010, *51*, 190–200.

[19] Toujani, W., Muñoz-Bertomeu, J., Flores-Tornero, M., Rosa-Téllez, S. et al., Identification of the phosphoglycerate dehydrogenase isoform EDA9 as the essential gene for embryo and male gametophyte development in *Arabidopsis*. *Plant Signal. Behav.* 2013, *8*, e27207.

[20] Truong, N.-H., Park, S.-M., Nishizawa, Y., Watanabe, T. et al., Structure, heterologous expression, and properties of rice (*Oryza sativa* L.) family 19 Chitinases. *Biosci. Biotechnol. Biochem.* 2003, *67*, 1063–1070.

[21] Jin, H., Song, Z., Nikolau, B. J., Reverse genetic characterization of two paralogous acetoacetyl CoA thiolase genes in Arabidopsis reveals their importance in plant growth and development. *Plant J.* 2012, *70*, 1015–1032.

[22] Fusaro, A. F., Bocca, S. N., Ramos, R. L. B., Barrôco, R. M. et al., AtGRP2, a cold-induced nucleo-cytoplasmic RNA-binding protein, has a role in flower and seed development. *Planta* 2007, *225*, 1339–1351.

[23] Yang, Y., Karlson, D. T., Overexpression of AtCSP4 affects late stages of embryo development in Arabidopsis. *J. Exp. Bot.* 2011, *62*, 2079–2091.

[24] Tonietto, Â., Sato, J. H., Teixeira, J. B., de Souza, E. M. et al., Proteomic analysis of developing somatic embryos of Coffea arabica. *Plant Mol. Biol. Report.* 2012, *30*, 1393–1399.

[25] Heringer, A. S., Barroso, T., Macedo, A. F., Santa-Catarina, C. et al., Label-free quantitative proteomics of embryogenic and non-embryogenic callus during sugarcane somatic embryogenesis. *PLoS One* 2015, *10*, e0127803.

[26] Varhaníková, M., Uvackova, L., Skultety, L., Pretova, A. et al., Comparative quantitative proteomic analysis of embryogenic and non-embryogenic calli in maize suggests the role of oxylipins in plant totipotency. *J. Proteomics* 2014, *104*, 57–65.

[27] Baba, A. I., Nogueira, F. C. S., Pinheiro, C. B., Brasil, J. N. et al., Proteome analysis of secondary somatic embryogenesis in cassava (Manihot esculenta). *Plant Sci.* 2008, *175*, 717–723.

